Semiannual Progress Report

Period Covered: 5/1/91-11/1/91

NASA Cooperative Research Agreement
NCC2-542

Psychophysical Evaluation of Three-Dimensional Auditory Displays

Frederic L. Wightman, Ph. D., Principal Investigator
Waisman Center, University of Wisconsin
1500 Highland Avenue
Madison, WI 53705

Summary Progress Report

Work during this reporting period included the completion of our research on the use of principal components analysis (PCA) to model the acoustical transfer functions (HRTFs) that are used to synthesize virtual sources for three dimensional auditory displays. In addition, a series of studies was initiated on the perceptual errors made by listeners when localizing free-field and virtual sources. Previous research has revealed that under certain conditions these perceptual errors, often called "confusions" or "reversals", are both large and frequent, thus seriously comprising the utility of a 3-D virtual auditory display. The long-range goal of our work in this area is to elucidate the sources of the confusions and to develop signal-processing strategies to reduce or eliminate them.

1. Completion of research on Principal Components Analysis of HRTFs:

HRTFs were measured from both ears of 10 subjects from 265 positions in an anechoic sound field. After the mean log-magnitude function was removed, the resulting log-magnitude functions were subjected to a principal components analysis. The analysis revealed that over 90% of the total variance in the 5300 HRTFs could be explained by 5 principal components. HRTFs were reconstructed by combining the PCA-derived log-magnitude functions with minimum-phase phase functions. Reconstructions of varying fidelity were obtained by using 1, 3, or 5 principal components, with the 5 principal component reconstructions providing the closest approximation to the original HRTFs. Subjects judged the apparent positions of sound sources synthesized from the reconstructed HRTFs. The 5-PC reconstruction resulted in localizations that were nearly as accurate as with stimuli synthesized from the original HRTFs. With fewer PCs used for the synthesis the frequency of azimuth and elevation confusions increased dramatically. The results of the PCA and the psychophysical experiments were described in a manuscript that was submitted and accepted for publication (included as Appendix to this report).

2. Studies of "confusion" errors:

Over the years, we have collected localization data from a large number of subjects in a variety of free-field and virtual free-field conditions. Errors that have been classified as "confusions" are evident in all of these data sets and thus constitute a rich resource for the study of such confusions. The purpose of the work begun during this reporting period was a thorough study of previous data, with a view toward identifying those features of the acoustical environment that lead to high and low rates of confusions.

The results reveal that confusion rates are always higher in virtual free-field conditions than in the free-field conditions they mimic and that rates are also higher with stimuli that have an uncertain spectrum from trial to trial. In addition, confusion rates increase as stimulus bandwidth is restricted. High-frequency content

seems especially important, in that confusion rates increased dramatically when high frequencies were removed by low-pass filtering.

It appears that any degradation of the spectral cues that are normally available to listeners, either by making the stimulus spectrum uncertain or by reducing the stimulus bandwidth causes increases in the rates of confusions. Thus the spectral cues seem especially important for resolving "cone-of-confusion" ambiguities that result from listeners' normal dependence on interaural time and intensity cues. The generally higher confusion rates in virtual free-field conditions could be a result either of errors in the synthesis that degrade the spectral cues or of the lack of the additional cues listeners receive in free-field from head movements.

Publications during reporting period:

Wightman, F., and Kistler, D. (1992). "The dominant role of low-frequency interaural time differences in sound localization", *Journal of the Acoustical Society of America*, in press.

Kistler, D. and Wightman, F. (1992). "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction", *Journal of the Acoustical Society of America*, in press.

Detailed Progress Report

1. Applicability of PCA for Modelling HRTFs and Synthesizing Virtual Sources:

       This work has been described in several previous progress reports, so need not be described again here. As mentioned above, the aim of our current efforts in this area was completion of the research and the publication of the results. The manuscript that resulted from this work was submitted for publication and accepted during this reporting period. A copy of the manuscript is attached to this report as Appendix A.

2. Analysis of "Confusion" Errors:

       One of the most troublesome characteristics of virtual sound sources is that the apparent spatial positions occasionally are far away from the intended positions. The most common manifestation of these large perceptual errors is that either the azimuth or elevation components of the intended position are reversed. For example, an azimuth reversal occurs when a source synthesized to appear in the front of the listener is perceived to be behind, or vice versa. An elevation reversal occurs when a source intended to appear above the horizontal plane that intersects the ears is judged to be below, or vice versa. Although listeners occasionally make such azimuth and elevation reversals (also called confusions) when localizing real sources, the rate is typically greater when localizing virtual sources. To insure the success of three-dimensional auditory displays, we feel that it is important to identify the sources of these confusions and to develop strategies for reducing their frequency. The first step is a detailed analysis of our existing data. Over the past five years, we have collected data from over 70 listeners in a variety of free-field and virtual source conditions, providing a rich database from which to extract patterns and trends.

       The first set of analyses was performed on the free-field database to identify the sources that are most likely to be confused in a free-field listening situation. Although we currently make measurements of HRTFs at 266 source positions (formerly 144), most listeners are tested on approximately 140 sources in our psychophysical paradigm. For the purpose of these analyses we examined the data of listeners who had been tested at least 6 times on a minimum of 100 sources (i.e., 6 repetitions for each source position). We included in the analyses only those source positions for which we have judgments from at least 8 listeners and for which we have a minimum of 100 judgments. Applying these criteria we retained a total of 41 listeners and 128 sources for analysis.

       Of the 41 listeners, 35 participated in the free-field experiments in which the stimulus was a wideband noise with a "scrambled" spectrum that was different on each trial. This stimulus was generated by dividing the spectrum into critical bands

4

and assigning a random intensity (uniform distribution, 20 dB range) to the noise spectrum level in each critical band. The purpose of the scrambled-spectrum stimulus was to prevent listeners from learning specific stimulus or transducer characteristics. In addition to these 35 listeners, there were eight (two of whom were in the scrambled spectrum experiments) who participated in a free-field experiment with a wideband noise stimulus, the spectrum of which was not scrambled. This stimulus allows listeners to take maximum advantage of the spectral cues provided by the pinna. It is also comparable to stimuli used by other researchers and thus enables a comparison of our data to the data collected in other laboratories.

For the purpose of our analyses, we identified reversals using the same liberal criterion as we had previously used (Wightman and Kistler, 1989, *JASA*, 85, p.872) in connection with resolving them before analysis of localization data. In short, we classified as a reversal any judgment for which the error (angular distance between judgment and target) could be reduced by reflecting the judgment about the lateral vertical plane (for front-back confusions) or the horizontal plane (for up-down confusions). This criterion admittedly overestimates the confusion rates, since for stimuli with an azimuth near plus/minus 90° or an elevation near 0° simple errors would be classified as confusions. However, we have also counted confusions with more strict (and more elaborate) criteria and found no substantial differences in the observed trends.

Figure 1 shows reversal rates plotted as a function of source position in the scrambled-spectrum free-field experiment. These rates were computed by summing reversals over listeners and dividing by the total number of trials for each position. Azimuth reversals are plotted in the top panel and elevation reversals in the bottom panel. The largest azimuth reversal rates occur for sources at high elevations, especially for sources in the front and for sources at 75° azimuth at all elevations. Neither result is surprising. Large changes in azimuth represent small changes in actual distance for sources at high elevations. Consequently reversal rates are confounded with potentially large response variability at high elevations. A similar explanation may hold for the response patterns for sources at 75°. Since 75° is only 15° from 90°, normal response variability may account for the high reversal rates here. However, it is surprising that equally high rates did not occur for sources at 105°. Elevation reversal rates were lower than azimuth reversals. Elevation reversals for sources near 0° elevation were most frequent. It is highly likely that some of these were not "true" reversals but normal response variability.

Table 1 shows the azimuth and elevation reversal rates for each listener. These rates were derived by summing the reversals across all source positions and representing the sum as a percentage of the total number of responses. For most listeners azimuth reversal rates are higher than elevation reversal rates. Of the 35 listeners, 20 made a significantly greater number of front/back reversals, and 4 made a significantly greater number of back/front reversals. The rates of the two types of azimuth reversals did not differ significantly for 4 listeners. For the remaining 7

listeners, azimuth reversal rates were less than 5% and could not be reliably compared. For most listeners a comparison of the two types of elevation reversals rates was not feasible since the rates were so low (i.e., under 5%). For the 12 listeners with rates greater than 5%, statistical tests indicated that the rate of down/up reversals was higher for 8, while the rate of up/down reversals was higher for 1 listener only.

In the non-scrambled-spectrum free-field experiment, 8 listeners were tested with 72 source positions. Figure 2 shows group azimuth and elevation reversal rates in the top and bottom panels, respectively. The rates for both types of reversals were lower than in the scrambled spectrum experiment. Azimuth reversals were greatest for sources at 75° and 105° and for sources at elevations greater than 45°. A greater percentage of back/front reversals occurred at the higher elevations in this experiment, while the reverse was true in the scrambled spectrum experiment. Although elevation reversals were infrequent, the rates were slightly higher for the low sources in the rear.

Table 2 provides the azimuth and elevation reversals rates for individual listeners. For the 5 listeners with azimuth reversal rates greater than 5%, 4 had a significantly greater number of back/front reversals, while 1 had more front/back reversals. We did not compare the two types of elevation reversals since the overall rates were so low. The higher reversal rates observed with the scrambled spectrum stimulus may indicate that any degradation of the spectral cues reduces the listeners ability to locate sources that are potentially confusable (e.g., sources on the same cone of confusion). However, it is also possible that smaller sample in the non-scrambled spectrum experiment was not representative of the population and that the lower rates are not realistic. Of the two listeners (SKT and SLN) who participated in both experiments, only SLN had significantly higher azimuth reversals in the scrambled spectrum experiment. It is also noteworthy that the reversal rates we observe in the non-scrambled spectrum experiment are close to those reported by others who have used non-scrambled-spectrum wideband stimuli.

The trends observed in the free-field experiments were also apparent in the virtual source experiments. Although the rates in the virtual source experiments were slightly higher than in the free-field experiments, the incidence patterns were very similar. Figure 3 shows the reversal rates as a function of intended source position for the scrambled spectrum experiment. These rates were computed from the data of 16 listeners, all of whom participated in the free-field experiment. The majority of the listeners (13) made significantly more front/back reversals. Eight of the listeners had elevation reversal rates greater than 5%. Of these, 3 had significantly more down/up reversals, 3 had more up/down reversals, and 2 showed no rate differences.

Figure 4 shows the reversal rates for the virtual source experiments that used the non-scrambled-spectrum stimulus. These rates are similar to those from the

comparable free-field experiment. The most notable differences are a small increase in front/back reversals at the high elevations and a small increase in up/down reversals. Reversal rates for the 6 listeners in this experiment are given in Table 4.

We have tested listeners in a number of experiments in which the HRTFs used to simulate the virtual sources were altered in some way. Thus far we have completed the reversal analyses for two series of experiments: the principal component experiments and the reduced bandwidth experiments. In the principal component experiments, which are described in the manuscript included as Appendix A, we noted an increase in both azimuth and elevation reversal rates as we decreased the number of principal components used to derive the HRTF, which was then used to produce the virtual sources. The reversal rates for the 36 "pc-derived" sources are plotted in Figures 5-7. The high front/back and back/front rates in the 3-component and 1-component conditions (Figures 6 and 7) are due to the fact that 3 of the 5 listeners judged most of sources to be in the front and 2 judged most to be in the back. Perhaps the most striking result is that all listeners made a large number of down/up elevation reversals in these "diminished" cue conditions. In the 1-component condition, most sources were judged to be above the listeners head.

In all the experiments discussed above, listeners were tested with a wideband stimulus (.2-14 kHz). We have also collected data on 8 listeners using virtual sources with stimuli that were filtered to reduce the bandwidth. Filtering was accomplished with a 10th order zero-phase FIR filter designed by the windowing method to approximate an ideal (infinite rejection rate above or below the cutoff frequency) highpass or lowpass filter. We have completed the reversal analyses of six of these experiments: two highpass conditions with cutoff frequencies of 5 kHz and 10 kHz, two lowpass conditions with cutoff frequencies of 5 kHz and 10 kHz, a bandpass condition with a lowpass cutoff of 5 kHz and a highpass cutoff of 10 kHz, and a bandstop condition in which all frequencies except those in the 5-10 kHz region were filtered out. We observed an increase in azimuth and elevation reversals in all of the reduced bandwidth conditions. The 10 kHz lowpass condition had the smallest increase in reversal rates relative to the baseline virtual source condition. The reversal rates for this condition are plotted in Figure 8. Although azimuth reversals increased for most locations, the sources in front of the listener at high elevation were most often confused, as was the case in the baseline condition (Figure 3). There was an increase in down/up reversals for sources in the front. All 8 listeners tended to "elevate" sources in the front.

The most notable result in the 5 kHz lowpass condition (Figure 9) was the increase in back/front reversals. Five of the 8 listeners made almost exclusively back/front reversals. These same listeners had made mostly front/back reversals in the baseline and 10 kHz lowpass conditions. Not surprisingly, elevation perception was diminished in this condition, since the 5-10 kHz region provides the primary spectral cues for elevation perception. Listeners judged the sources in the front at low elevations to be high and the sources in the back at high elevations to be low.

In the bandstop condition (Figure 10), which was similar to the 5 kHz lowpass condition, except the frequencies above 10 kHz were present, the tendency to elevate sources in the front disappeared and the front/back reversals again dominated the performance of all listeners.

In the 5 kHz highpass condition (Figure 11), listeners made more azimuth reversals, primarily front/back, than in the baseline condition. Elevation perception was only slightly affected. In the bandpass condition (Figure 12), the stimulus contained information in the 5-10 kHz region only. Azimuth performance was similar to the 5 kHz highpass condition in which all listeners showed an increase in front/back reversals relative to the baseline condition. However elevation reversals increased in this condition. There was an increase in down/up reversals at all azimuths. Thus with the frequencies above 10 kHz missing, listeners tended to overestimate the elevation of sources. Both azimuth and elevation perception was dramatically reduced in the 10 kHz highpass condition (Figure 13). Virtually all source locations were judged to be in the back and low.

The results of these preliminary analyses suggest that the spectral cues provided across the entire spectrum are important for resolving the "cone-of-confusion" ambiguities that produce front/back and up/down confusions. Work on more refined analyses and definition of strategies for reducing confusions in virtual source conditions is in progress.

TABLE 1. Percentage of Azimuth and Elevation reversals in the scramble-spectrum free-field experiment.

ID	Azimuth	Elevation	ID	Azimuth	Elevation
SDE	15.8	25.8[3]	SHF	11.7[1]	4.3
SDH	7.8[1]	3.1	SHG	13.2[1]	3.8
SDL	8.8[1]	2.8	SIK	3.8	1.2
SDM	7.6[1]	2.5	SIO	5.3[2]	7.3[3]
SDO	4.8	1.0	SIP	15.6[1]	6.1
SDP	3.7	1.7	SIS	7.9[1]	5.1[3]
SEB	3.8	0.4	SJX	4.8	1.4
SED	4.6	0.9	SKG	15.2[1]	2.5
SER	6.3[1]	1.9	SKH	24.0[1]	7.9[3]
SES	14.7[1]	11.1[3]	SKT	10.3[2]	5.1[3]
SET	9.1[1]	0.8	SLN	11.5[2]	1.4
SFI	11.8	0.8	SLT	5.5[1]	4.3
SGB	7.6[1]	1.5	SLU	12.9[2]	8.3
SGC	3.8	1.2	SLV	10.2[1]	5.5[3]
SGD	5.7	6.6	SLW	9.9[1]	4.7
SGE	15.4[1]	2.7	SLX	9.3[1]	8.8[3]
SGG	6.7[1]	2.2	SLZ	19.3[1]	16.0[4]
SHD	11.9	1.6			

[1] Front/back reversals were significantly greater than back/front reversals.
[2] Back/front reversals were significantly greater than front/back reversals.
[3] Down/up reversals were significantly greater than up/down reversals.
[4] Up/down reversals were significantly greater than down/up reversals.

TABLE 2. Percentage of azimuth and elevation reversals in the non-scrambled-spectrum free-field experiment.

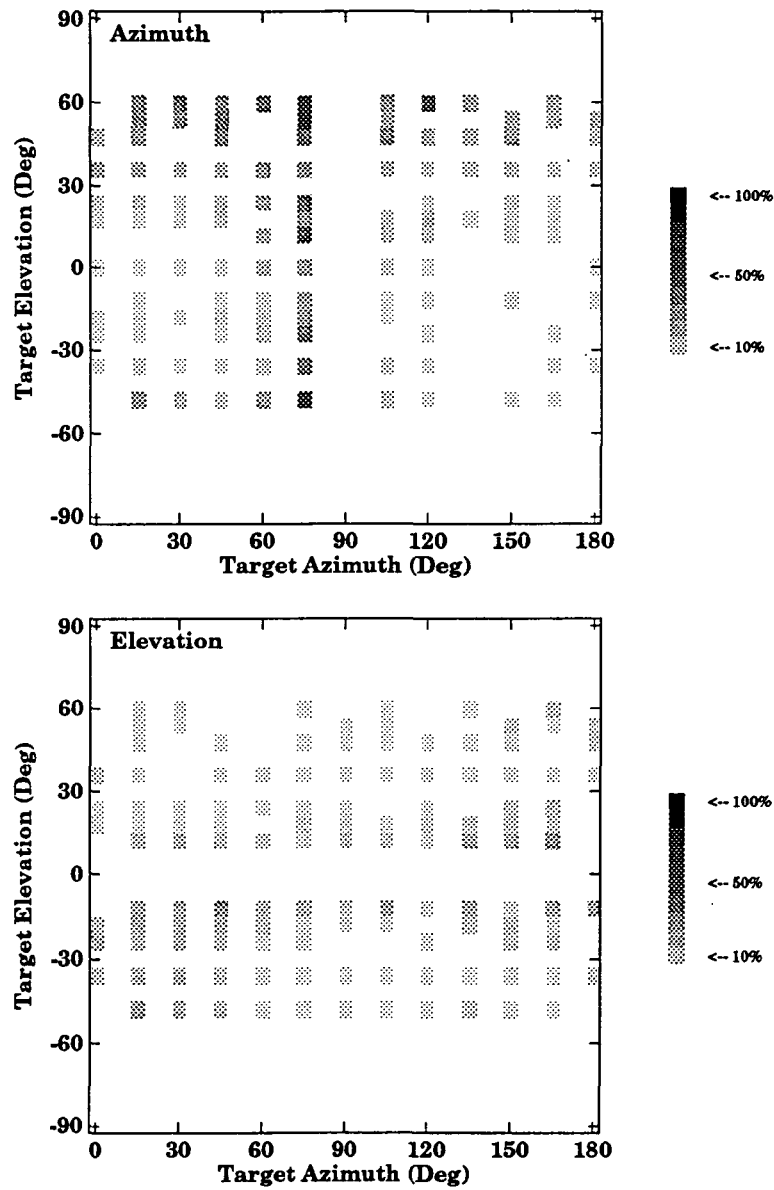| ID	Azimuth	Elevation
SKR	6.1^2	3.5
SKS	3.2	1.8
SKT	9.8^2	2.9
SLE	1.8	2.2
SLG	3.8	1.0
SLN	6.0^2	1.5
SLO	5.3^1	2.2
SLQ	5.3^2	2.3

[1] Front/back reversals were significantly greater than back/front reversals.
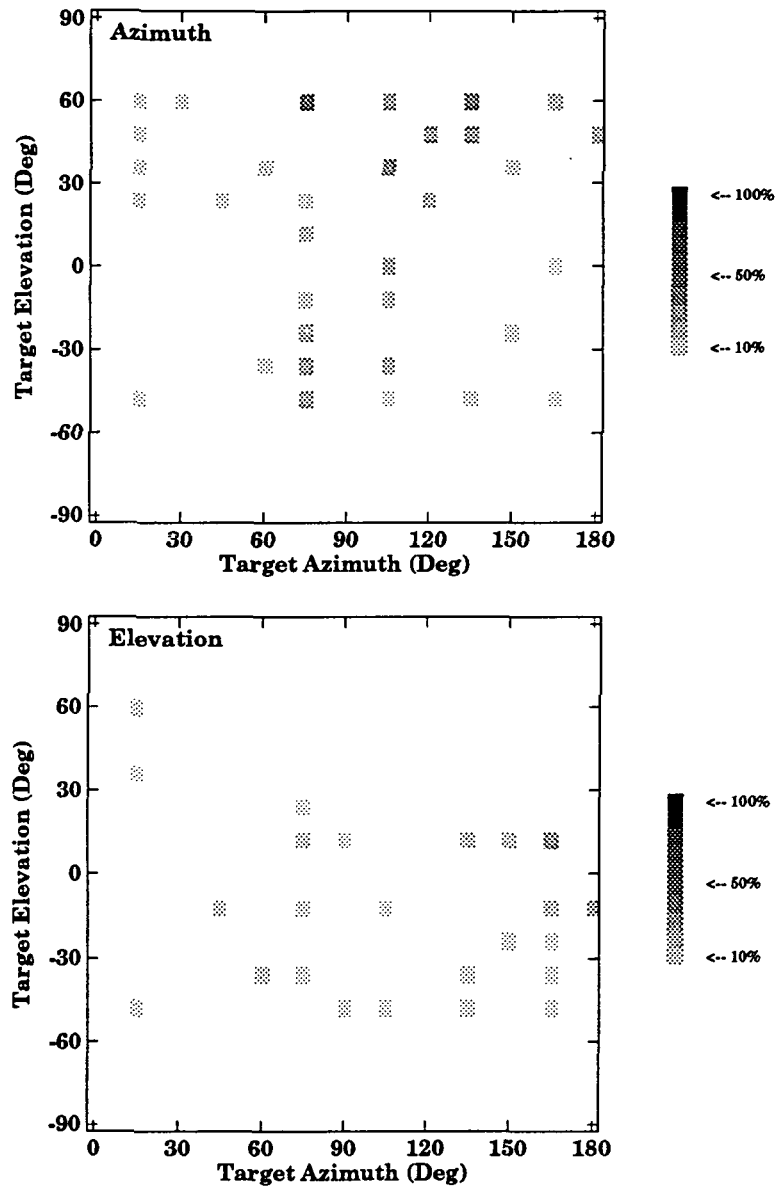
[2] Back/front reversals were significantly greater than front/back reversals.

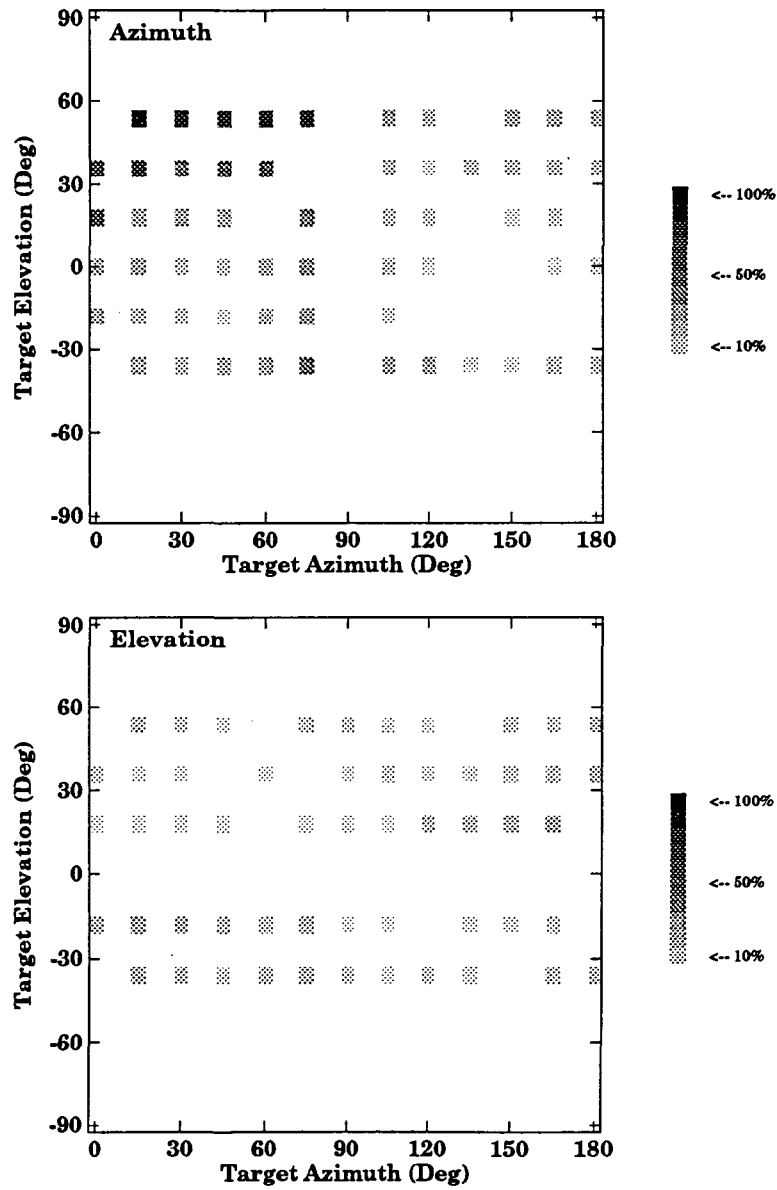[3] Down/up reversals were significantly greater than up/down reversals.

[4] Up/down reversals were significantly greater than down/up reversals.

TABLE 3. Percentage of azimuth and elevation reversals in the scrambled-spectrum virtual source experiment.

ID	Azimuth	Elevation
SDE	23.0^1	33.3^3
SDH	13.5^1	3.0
SDL	18.3^1	4.0
SDM	10.6	7.3
SDO	12.5^1	1.8
SDP	8.7^1	2.8
SED	7.1^1	5.0
SER	9.2^2	2.3
SET	19.6^1	7.8^4
SGB	33.2^1	4.3
SGD	11.1^1	12.4^4
SGE	23.8^1	4.9
SGG	10.6^1	5.4^3
SHD	14.3^1	3.0
SHG	15.3	6.6^4
SIK	27.6^1	12.4^3

[1] Front/back reversals were significantly greater than back/front reversals.
[2] Back/front reversals were significantly greater than front/back reversals.
[3] Down/up reversals were significantly greater than up/down reversals.
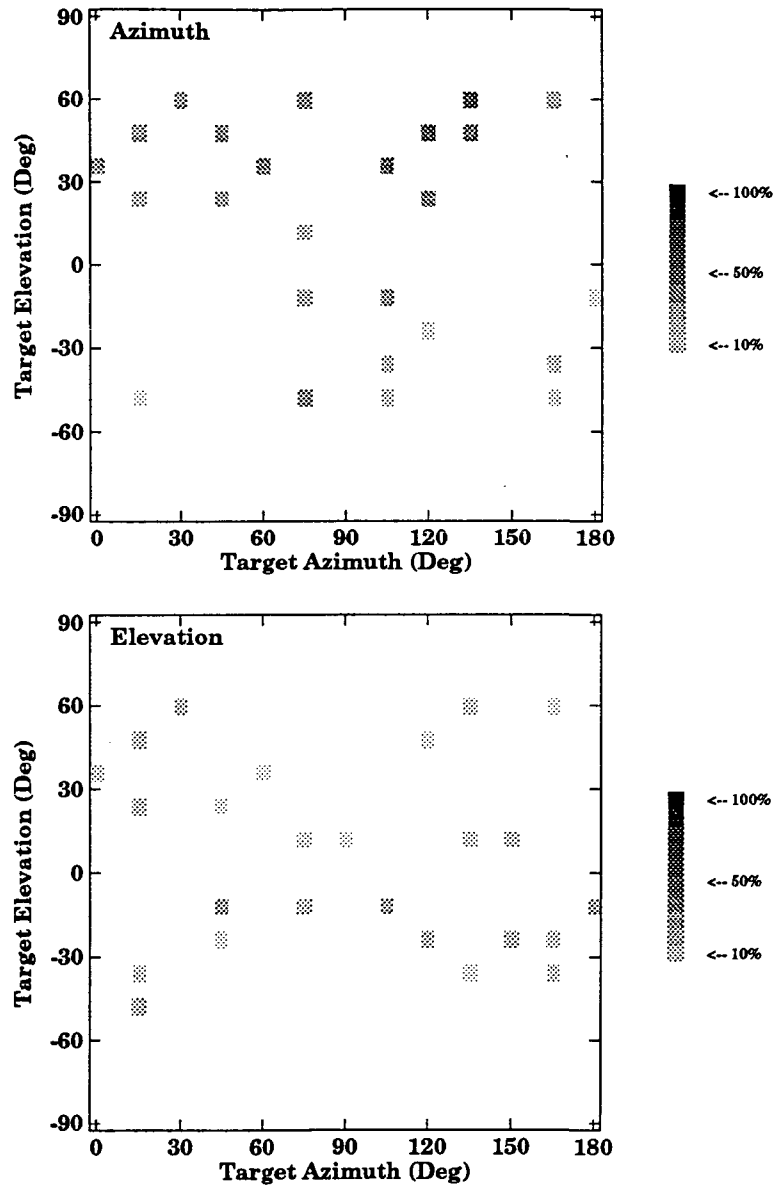[4] Up/down reversals were significantly greater than down/up reversals.

TABLE 4. Percentage of azimuth and elevation reversals in the non-scrambled-spectrum virtual source experiment.

ID	Azimuth	Elevation
SKR	10.3^2	6.1^3
SKS	10.5^1	5.7^4
SKT	22.3^1	2.1
SLG	9.0^2	1.5
SLN	2.6	.0
SLO	9.6	9.2^3

[1] Front/back reversals were significantly greater than back/front reversals.

[2] Back/front reversals were significantly greater than front/back reversals.

[3] Down/up reversals were significantly greater than up/down reversals.

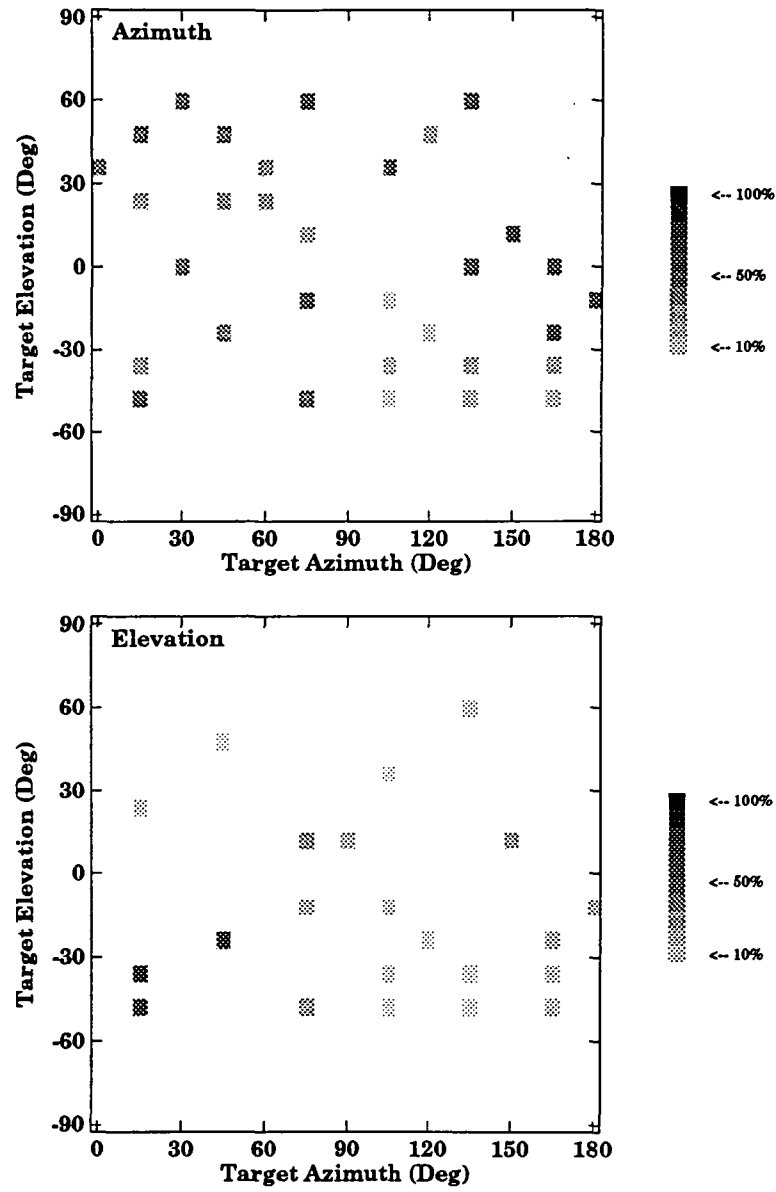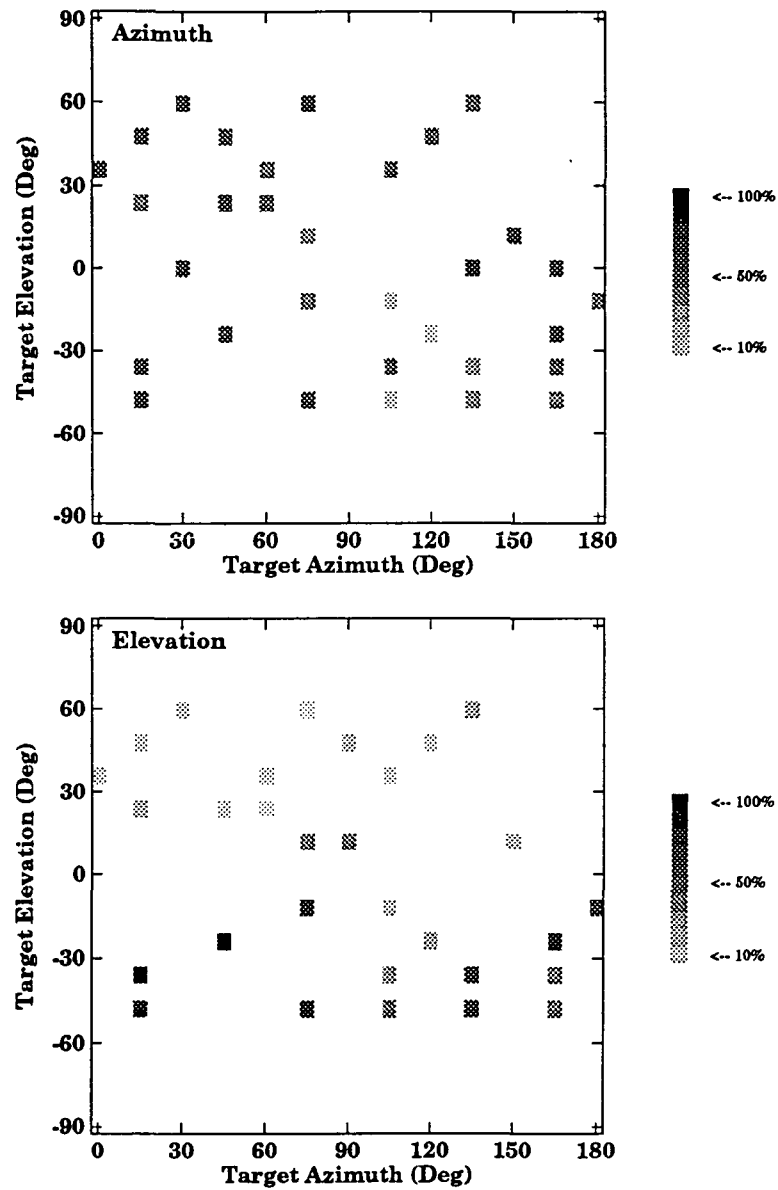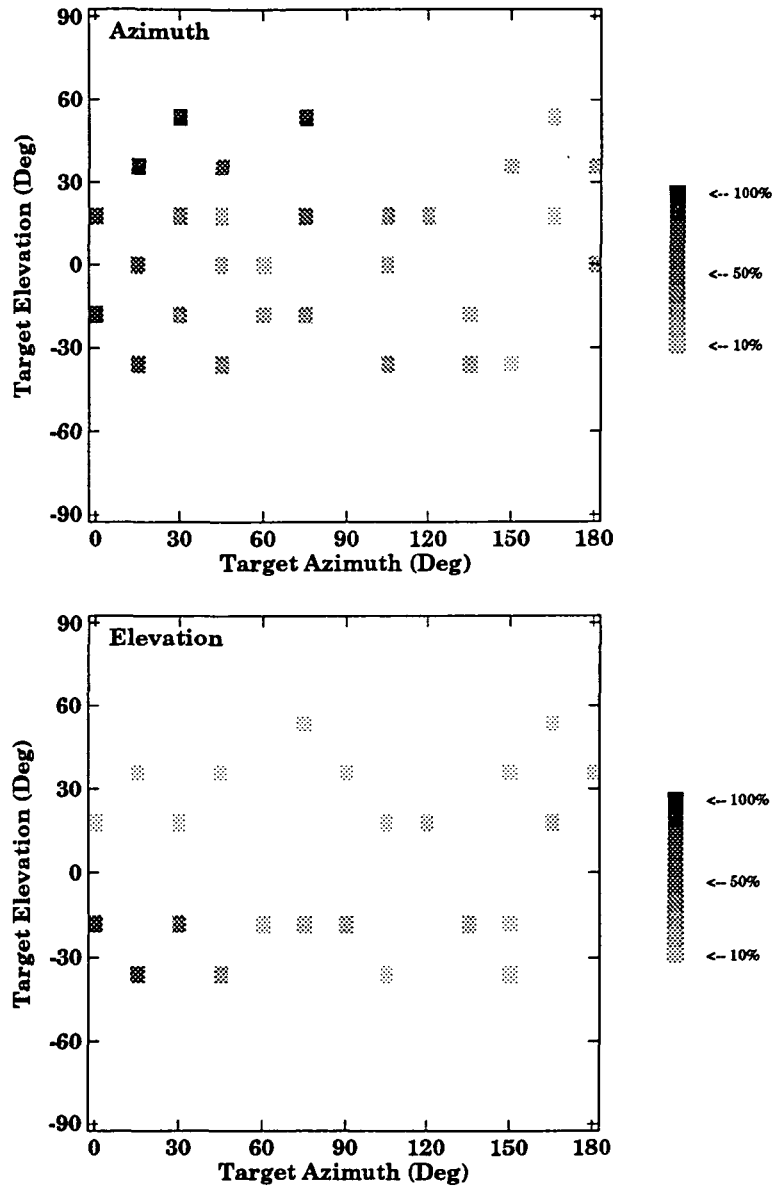[4] Up/down reversals were significantly greater than down/up reversals.

Figure 1. The percentage of azimuth (top panel) and elevation (bottom panel) reversals plotted as a function of target azimuth and elevation for the scrambled-spectrum free-field experiment.

13

Figure 2. *The percentage of azimuth (top panel) and elevation (bottom panel)* reversals plotted as a function of target azimuth and elevation for the non-scrambled-spectrum free-field experiment.

14

Figure 3. The percentage of azimuth (top panel) and elevation (bottom panel) reversals plotted as a function of target azimuth and elevation for the scramble-spectrum virtual source experiment.

Figure 4. The percentage of azimuth (top panel) and elevation (bottom panel) reversals plotted as a function of target azimuth and elevation for the non-scrambled-spectrum virtual source experiment.

16

Figure 5. The percentage of azimuth (top panel) and elevation (bottom panel) reversals plotted as a function of target azimuth and elevation for the 5-component condition of the PCA experiment.
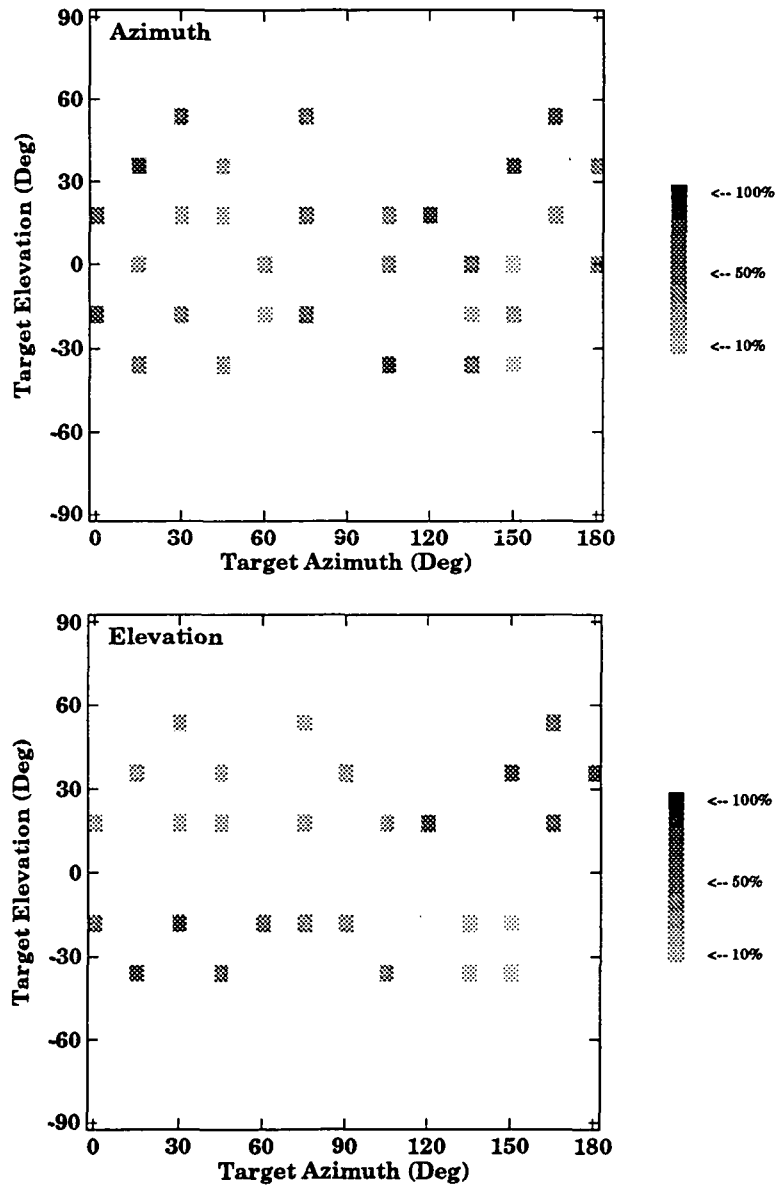
Figure 6. The percentage of azimuth (top panel) and elevation (bottom panel) reversals plotted as a function of target azimuth and elevation for the 3-component condition of the PCA experiment.

18

Figure 7. The percentage of azimuth (top panel) and elevation (bottom panel) reversals plotted as a function of target azimuth and elevation for the 1-component condition of the PCA experiment.

Figure 8. The percentage of azimuth (top panel) and elevation (bottom panel) reversals plotted as a function of target azimuth and elevation for the 10 kHz lowpass filter condition of the reduced bandwith experiment.
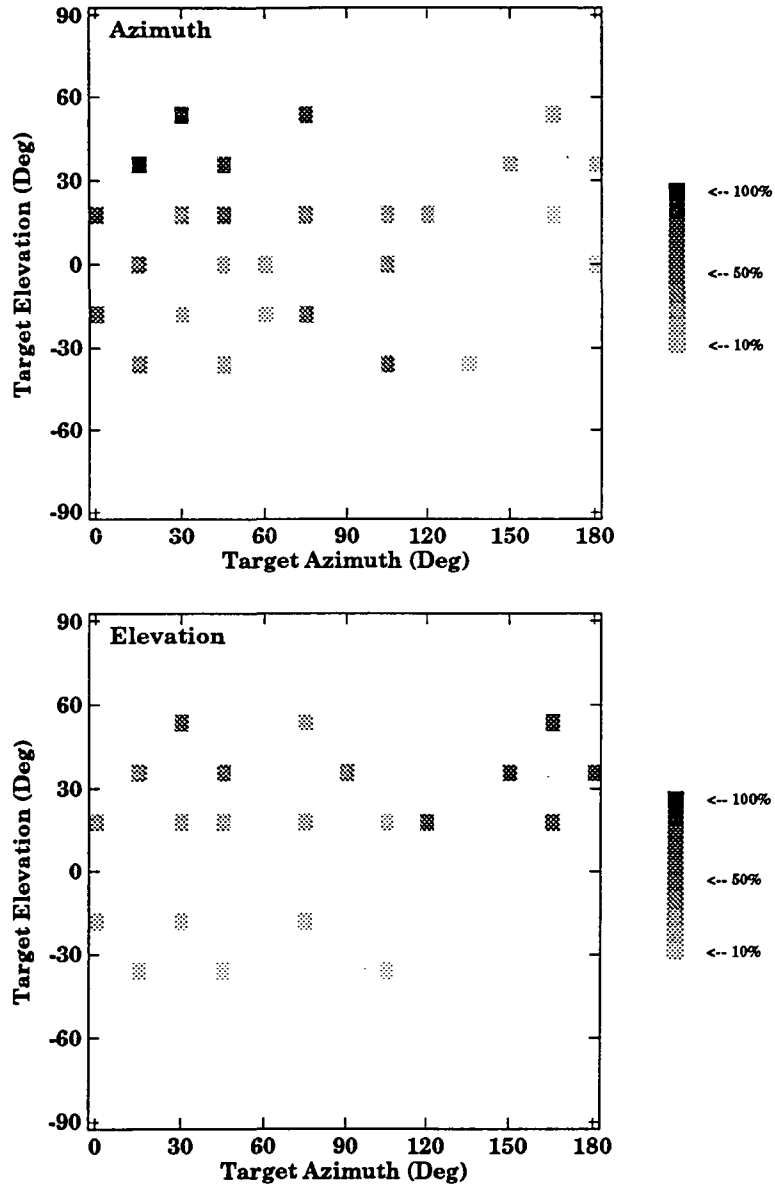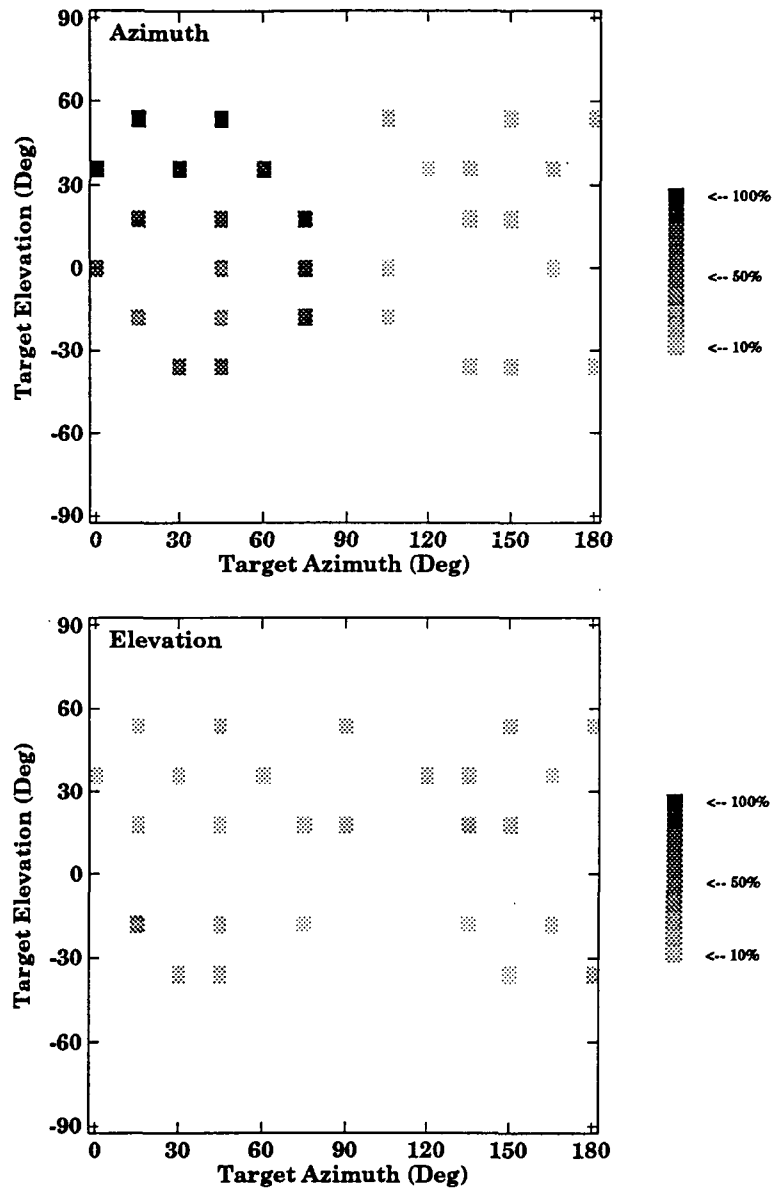
20

Figure 9. The percentage of azimuth (top panel) and elevation (bottom panel) reversals plotted as a function of target azimuth and elevation for the 5 kHz lowpass filter condition of the reduced bandwith experiment.
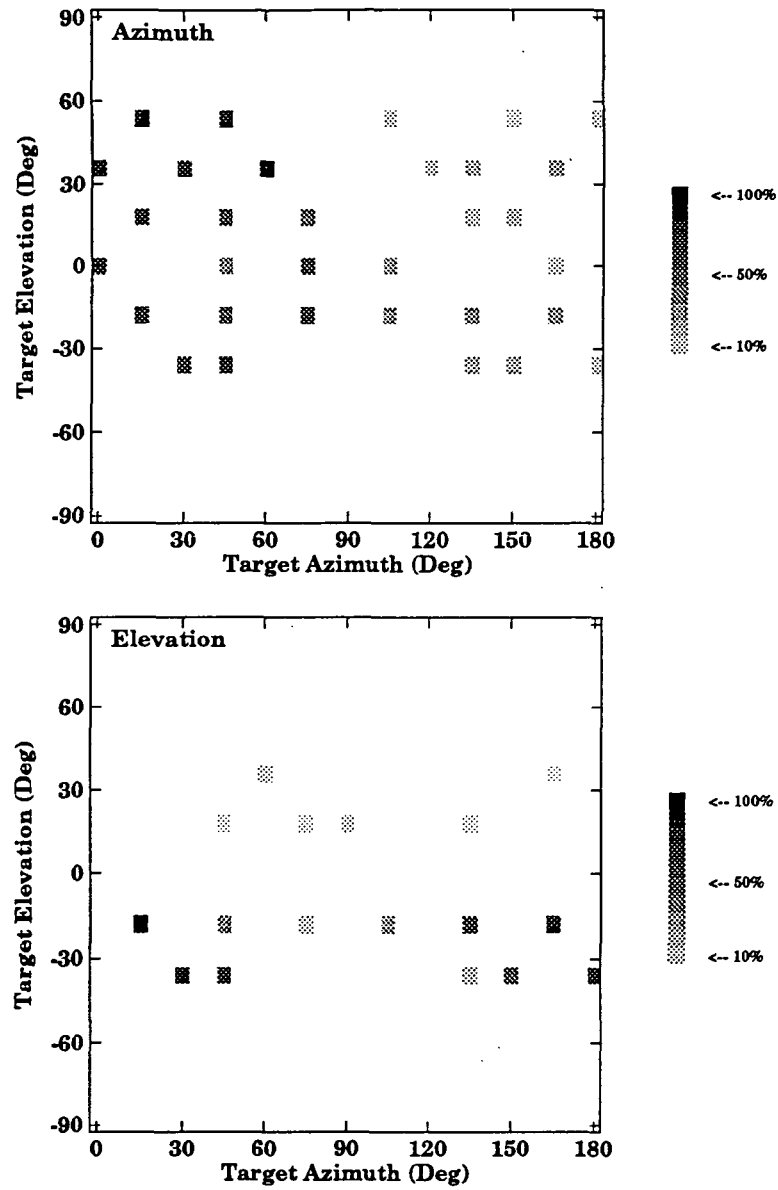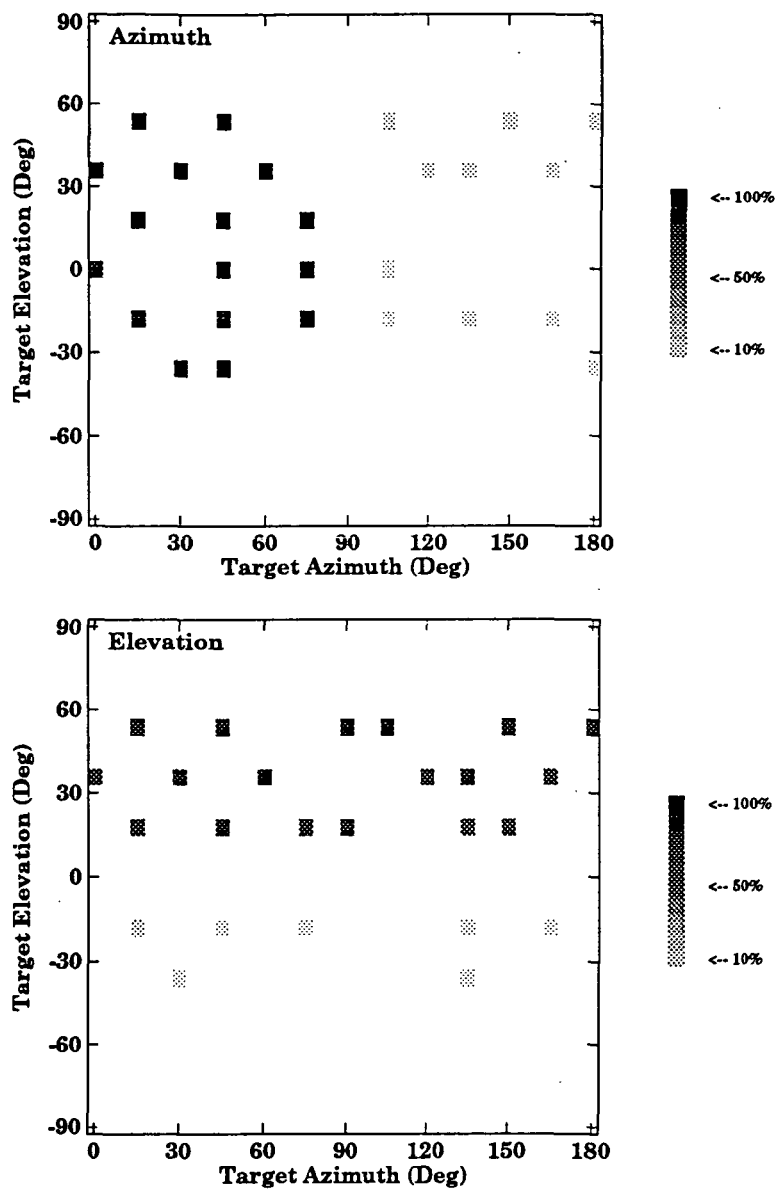
Figure 10. The percentage of azimuth (top panel) and elevation (bottom panel) reversals plotted as a function of target azimuth and elevation for the bandstop filter condition of the reduced bandwith experiment.

Figure 11. The percentage of azimuth (top panel) and elevation (bottom panel) reversals plotted as a function of target azimuth and elevation for the 5 kHz highpass filter condition of the reduced bandwidth experiment.

Figure 12. The percentage of azimuth (top panel) and elevation (bottom panel) reversals plotted as a function of target azimuth and elevation for the bandpass filter condition of the reduced bandwidth experiment.

Figure 13. The percentage of azimuth (top panel) and elevation (bottom panel) reversals plotted as a function of target azimuth and elevation for the 10 kHz highpass filter condition of the reduced bandwidth experiment.